



# Optimization of the Hölder Image Descriptor using a Genetic Algorithm

Leonardo Trujillo, Pierrick Legrand, Gustavo Olague, Cynthia Pérez

## ► To cite this version:

Leonardo Trujillo, Pierrick Legrand, Gustavo Olague, Cynthia Pérez. Optimization of the Hölder Image Descriptor using a Genetic Algorithm. GECCO 2010. Best paper award in "Real world applications", Jul 2010, Portland Oregon, United States. pp.1147-1154. inria-00534457

**HAL Id: inria-00534457**

**<https://hal.inria.fr/inria-00534457>**

Submitted on 28 Nov 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimization of the Hölder Image Descriptor using a Genetic Algorithm

Leonardo Trujillo  
Instituto Tecnológico de  
Tijuana  
Computer Science Graduate  
Program  
Tijuana, B.C., México  
leonardo.trujillo.ttl@gmail.com

Pierrick Legrand  
Université Victor Segalen  
Bordeaux 2  
ALEA Team, INRIA Bordeaux  
Sud-Ouest  
IMB, UMR CNRS 5251,  
France.  
pierrick.legrand@u-  
bordeaux2.fr

Gustavo Olague,  
Cynthia Pérez  
EvoVisión Project  
CICESE Research Center  
Ensenada, B.C. México  
olague@cicese.mx  
cbperez@cicese.mx

## ABSTRACT

Local image features can provide the basis for robust and invariant recognition of objects and scenes. Therefore, compact and distinctive representations of local shape and appearance has become invaluable in modern computer vision. In this work, we study a local descriptor based on the Hölder exponent, a measure of signal regularity. The proposal is to find an optimal number of dimensions for the descriptor using a genetic algorithm (GA). To guide the GA search, fitness is computed based on the performance of the descriptor when applied to standard region matching problems. This criterion is quantified using the F-Measure, derived from recall and precision analysis. Results show that it is possible to reduce the size of the canonical Hölder descriptor without degrading the quality of its performance. In fact, the best descriptor found through the GA search is nearly 70% smaller and achieves similar performance on standard tests.

## Categories and Subject Descriptors

I.4.7 [Image Processing and Computer Vision]: Feature Measurement—*invariants*; I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search—*heuristic methods*

## General Terms

Algorithms, Experimentation, Performance

## Keywords

Hölderian regularity, image descriptors, genetic algorithms

## 1. INTRODUCTION

Currently, a large part of computer vision research is devoted towards the development of recognition systems that

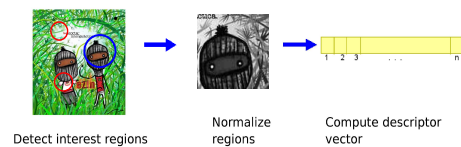


Figure 1: Detection/description of local image features.

rely on the analysis of salient local features. These features are commonly called interest points or interest regions [16, 20, 21]. This local approach has gained a wide acceptance because it can help reduce the severity of several practical problems. For instance, it is less sensitive to partial occlusions within the scene [11], it does not require traditional image segmentation [15], and provides a higher invariance to geometric and photometric transformations [16, 13]. Furthermore, this analysis is conceptually simple and can be easily adapted to different problem domains.

The basic approach consists of two phases, detection and description of locally salient features, see Figure 1. First, an operator identifies the position and scale of the salient image features [26]. Afterwards, each region is normalized, adjusting for scale, rotation and illumination invariance. Finally, each normalized region is given as input to the description process, which then outputs a numerical vector called a local descriptor [13]. These descriptors extract a compact and unique representation of local image structure, they are required to be distinctive and informative. It is clear that the performance of a system that uses this approach will depend on the performance of the algorithms that are used for detection and description of the salient regions.

Keeping to the problem of region description, many proposals have been developed [13]. Probably the most widely used method is the Scale Invariant Feature Transform (SIFT) [11], another example is the more recent Hölder descriptor [24]. In both cases, a measure of signal variation is used to build an histogram that characterizes the local shape and appearance; the former relies on the gradient orientation, and the latter on a measure of pointwise regularity. The performance of both descriptors has been shown to be quite similar when applied on standard tests [24]. On the other hand, one drawback shared by both descriptors is that current implementations are relatively slow, especially when considering

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO '10, July 7–11, 2010, Portland, Oregon, USA.

Copyright 2010 ACM 978-1-4503-0072-8/10/07 ...\$10.00.

real-time applications. However, one important practical difference between the two is that SIFT uses a very elaborate algorithm that is not easy to reproduce, while the Hölder descriptor employs a much simpler and direct algorithm. As such, the latter is much more amenable to optimization.

Therefore, the goal of the present work is to develop an optimized version of the Hölder descriptor that might lead to a simpler description process without decreased performance. This goal is posed as a combinatorial search problem, in order to find the optimal number of dimensions for the Hölder descriptor. It is hypothesized that the optimal size of the descriptor might be smaller than the original proposal. This hypothesis is based on the assumption that a local descriptor might be redundant, and empirical evidence supports this claim [6]. If this is true for the Hölder descriptor, it could lead to a more compact description of local image features.

In order to achieve the goal stated above, we propose to use a genetic algorithm (GA) because of the combinatorial nature of the problem and the limited knowledge of the search space. Fitness is assigned based on the number of correct matches that are computed between two images using the descriptor. This criterion is quantified using an analysis of recall and precision statistics with the F-Measure, following the work of [14]. The evolutionary process could eliminate redundant and unnecessary dimensions if such a representation achieves optimal performance, thereby compressing image information even further. As such, the current proposal is closely related with a long list of GA-based methods developed for dimensionality reduction, see for example [17, 1, 18, 4, 23]. The results presented in this work show that the GA is indeed capable of finding a smaller local descriptor that achieves a similar performance.

The remainder of this paper proceeds as follows. Section 2 reviews the topic of local image descriptors, explains how descriptors can be evaluated [13], and outlines how they can be optimized [14]. Section 3 introduces the concept of Hölderian regularity and describes the canonical Hölder descriptor. Then, Section 4 describes the problem of this work and presents the proposed solution using a GA. Details of our implementation and the experiments are presented in Section 5. Finally, Section 6 contains our concluding remarks.

## 2. LOCAL IMAGE DESCRIPTORS

This section presents a brief overview of the state-of-the-art in local image descriptors, and describes a common approach for evaluation and comparison.

### 2.1 Previous work

Computer vision literature that focuses on the detection and description of local features has grown rapidly over the last ten years. A comprehensive review on these topics can be found in [26] for detection algorithms, and in [13] for description methods. Keeping to the latter, it is possible to identify four main groups of methods: distribution-based, spatial-frequency techniques, differential descriptors, and others. In this discussion we will only deal with the first group, because they have shown to be better at extracting distinctive image information [13]. Distribution-based descriptors use histograms to represent local image shape or appearance. Currently it is widely accepted that the SIFT descriptor is the best histogram based method, and it is probably the most widely used approach in computer vision research. The SIFT descriptor is a 3D histogram of gradient locations and orien-

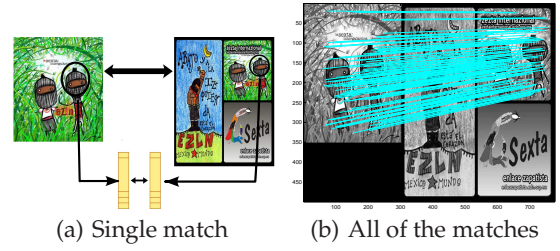


Figure 2: The matching process with local descriptors.

tations, where the contribution to each bin in the histogram is weighted by the gradient magnitude.

The success of SIFT has prompted many researchers to propose variations and improvements over the basic SIFT method, for examples see [13, 6, 2]. However, despite its success, there is one practical drawback to SIFT, it is a complex and relatively slow algorithm. Therefore, it is not a simple task to reproduce the original code, and it is not feasible to use SIFT in demanding applications that require real-time output [3]. Therefore, several researchers have proposed to reduce the dimensions of the SIFT vector [6] or to use simplified implementations [2].

### 2.2 Evaluation method

Let us now return to the topic of establishing an overall performance measure for local image descriptors. In this respect we follow [13], that bases the evaluation on recall and 1-precision curves. The problem used for evaluation is the matching of local regions between two different images of the same scene; see Figure 2.

For example, region **A** from image  $I_1$  is matched with region **B** from image  $I_2$  if the Euclidean distance between their corresponding descriptors,  $D_A$  and  $D_B$ , is below a certain threshold  $h$ , and if  $D_B$  is the nearest neighbor of  $D_A$ . When the geometric transformation between  $I_1$  and  $I_2$  is known beforehand, then it is possible to determine if each match is correct [13]. From this, recall and 1-precision values can be easily obtained using

$$\text{recall} = \frac{\# \text{ correct matches}}{\# \text{ true correspondences}},$$

$$1 - \text{precision} = \frac{\# \text{ false matches}}{\# \text{ correct matches} + \# \text{ false matches}}.$$

A performance curve for a descriptor can be built by varying the matching threshold  $h$ . In this work, we use twenty different values, following [13].

However, evaluating descriptors in this way does have some limitations. More notably, it causes ambiguities when the curves of two different descriptors intersect. In order to simplify the comparison between two curves we could use the F-Measure, as done in [14]. The F-Measure is a concept commonly used in information retrieval, it gives an estimation of the accuracy of a test; a perfect accuracy would produce an F-Measure equal to one, and zero in the opposite case. The F-Measure is defined as,

$$F(\text{precision}, \text{recall})_\beta = \frac{(1 + \beta^2) \cdot (\text{precision} \cdot \text{recall})}{\beta^2 \cdot \text{precision} + \text{recall}}, \quad (1)$$

where if  $\beta = 1$  we obtain a symmetric balance between precision and recall.

### 2.3 Optimization of the detection/description methods

The large number of proposed methods for detecting and describing locally salient features has necessitated the development of experimental evaluation methods such as the one described above. The goal of such measures of performance was to provide objective criteria that could be used to make an informed decision when choosing a method for a particular vision application. On the other hand, these measures have also facilitated the development of automatic design algorithms that automatically synthesize detection and description methods. For example, [16] proposed a performance measure for interest point detection based on point repeatability. Afterwards, this measure has been used to pose single [20, 22] and multi-objective [25] optimization problems that search for optimal interest point detectors. The evolved detectors achieve state-of-the-art performance, in other words they are human competitive results. Another example is the proposal made in [14], that uses the evaluation method described above to synthesize a novel weight operator for the SIFT algorithm. Whereas SIFT uses the gradient magnitude to weigh the contributions made to each bin in the histogram, the weight operators evolved with GP produced significant performance gains. It is of interest to note that previous proposals to enhance the SIFT descriptor failed to notice that the weighting function could, or should, be improved. Indeed, the original proposal seems reasonable to a human expert; fortunately, however, evolution need not be hindered by such *reasonable assumptions*, evidenced by the counter-intuitive operators found by the GP search [14]. From this it follows that further inquiry is still necessary regarding the development of new local descriptors. In this sense, we argue that the Hölder descriptor could prove to be a viable alternative for future work in this area [24].

## 3. THE HÖLDER DESCRIPTOR

In this section the concepts of local regularity and the Hölder exponent are introduced, and then the Hölder-based descriptor is described in detail.

### 3.1 Holderian regularity

It is known that most of the useful information contained within a signal is located within the irregular or singular regions. In images, for instance, such regions correspond with edges, corners and interest points. Hölderian regularity provides a characterization of such singular structures [12]. It can be quantified, for example, by the pointwise Hölder exponent which is defined as follows.

**Definition 1:** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $s \in \mathbb{R}^{+*} \setminus \mathbb{N}$  and  $x_0 \in \mathbb{R}$ .  $f \in C^s(x_0)$  if and only if  $\exists \eta \in \mathbb{R}^{+*}$ , and a polynomial  $P$  of degree  $< s$  and a constant  $c$  such that

$$\forall x \in B(x_0, \eta), |f(x) - P(x - x_0)| \leq c|x - x_0|^s, \quad (2)$$

where  $B(x_0, \eta)$  is the local neighborhood around  $x_0$  with a radius  $\eta$ . The pointwise Hölder exponent of  $f$  at  $x_0$  is  $\alpha_p(x_0) = \sup_s \{f \in C^s(x_0)\}$ .

Hölderian regularity refines the concept of the Taylor series approximation of a function by also accounting for non-differentiable points [12]. The pointwise Hölder exponent has proven to be useful in several tasks of image analysis, such as noise removal [9], interpolation [8], and edge detection [10]. However, it can only be computed analytically for a

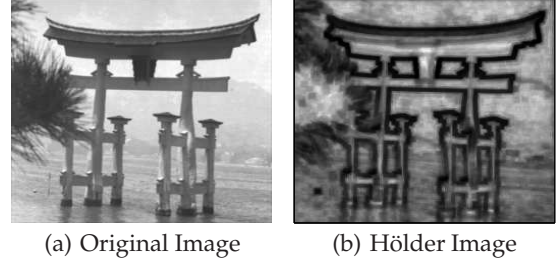


Figure 3: The pointwise Hölder exponent.

small set of signals. Therefore, in order to use Hölderian regularity the exponents must be estimated. Here we review the oscillations method for estimation, which is directly derived from the definition given above [19].

#### 3.1.1 Estimation through oscillations

The Hölder exponent of function  $f(t)$  at  $t$  is the  $\sup(\alpha_p) \in [0, 1]$ , for which a constant  $c$  exists such that  $\forall t'$  in a neighborhood of  $t$ ,

$$|f(t) - f(t')| \leq c|t - t'|^{\alpha_p}. \quad (3)$$

In terms of signal oscillations, a function  $f(t)$  is Hölderian with exponent  $\alpha_p \in [0, 1]$  at  $t$  if  $\exists c \forall \tau$  such that  $osc_\tau(t) \leq c\tau^{\alpha_p}$ , with

$$osc_\tau(t) = \sup_{t', t'' \in [t-\tau, t+\tau]} |f(t') - f(t'')|. \quad (4)$$

Now, if  $t = x_0$  and  $t' = x_0 + h$  in 3, we can also write that

$$\alpha_p(x_0) = \liminf_{h \rightarrow 0} \frac{\log |f(x_0 + h) - f(x_0)|}{\log |h|}. \quad (5)$$

Therefore, the problem is that of finding an  $\alpha_p$  that satisfies 3 and 4, and in order to simplify this process we can set  $\tau = \beta^r$ . Then, we can write  $osc_\tau \approx c\tau^{\alpha_p} = \beta^{(\alpha_p r + b)}$ , which is equivalent to  $\log_\beta(osc_\tau) \approx \alpha_p r + b$ .

An estimation of the regularity can be built at each point by computing the slope of the regression between the logarithm of the oscillations  $osc_\tau$  and the logarithm of the dimension of the neighborhood at which the oscillations  $\tau$  are computed; we use least squares regression with  $\beta = 2$  and  $r = 1, 2, \dots, 7$ . Also, it is preferable not to use all sizes of neighborhoods between two values  $\tau_{min}$  and  $\tau_{max}$ . Hence, we calculate the oscillation at point  $x_0$  only on intervals of the form  $[x_0 - \tau_r : x_0 + \tau_r]$ . For a 2D signal,  $x_0$  defines a point in 2D space and  $\tau_r$  a radius around  $x_0$ , such that  $d(t', t) \leq \tau_r$  and  $d(t'', t) \leq \tau_r$ , where  $d(a, b)$  is the Euclidean distance between  $a$  and  $b$ . Figure 3 presents the estimation of the Hölder exponent for an image. This method has proven to be superior in some cases to the wavelet leaders method [7, 5], and it is used to construct the Hölder descriptor we introduce below.

### 3.2 Hölder descriptor

The descriptor based on Hölderian regularity is very simple and easy to construct [24]. The idea is to uniformly sample the value of the Hölder exponent using a circular grid within each region. For instance, Figure 4a shows an interest point detected within a test image, and Figure 4b presents the local region around it. The descriptor is then constructed by



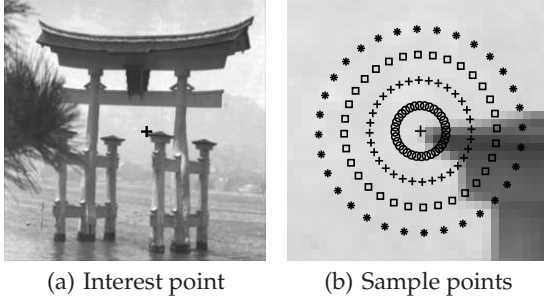


Figure 4: Sampling used with the Hölder descriptor.

sampling the exponent of the central point and at 32 equidistant points at four different radii, this gives a vector dimension of 129, Figure 4b illustrates this process. The Hölder descriptor has two useful properties for region description. First, because the exponent is estimated using oscillations which are relative intensity differences within the region, there is no need to normalize the descriptor for uniform intensity variations. Second, rotation invariance can be obtained by ordering the values in the descriptor based on the principal orientation of the gradient within each region. Additionally, it compares favorably with SIFT in two important aspects. On the one hand, experimental results have confirmed that the Hölder descriptor can achieve comparable results on standard tests [24]. On the other hand, the Hölder descriptor is constructed using a much simpler algorithm, this makes it easy to implement and replicate.

The biggest drawback of the Hölder descriptor is the high computation time it requires. Similar to SIFT, it is not feasible to use the Hölder descriptor in real-time. However, unlike SIFT, because the algorithm used to build the descriptor is very simple, there are two obvious ways in which to speed-up the process. One option is to devise a faster estimation method, however this is not a trivial task and is left for future work. Another option is to reduce the number of sample points used to build the descriptor. In [6] the dimensions of the SIFT descriptor were reduced through PCA, and they showed that some of the dimensions were not necessary to uniquely describe a region. Similarly, we suggest that this is a real possibility for the Hölder descriptor, and we expect that the optimal number of sample points might not be the 129 points used by the canonical version of the descriptor.

#### 4. THE SEARCH PROBLEM AND THE PROPOSED SOLUTION

Given the above arguments, the goal of this work is to find the optimal set of sample points that should be used to build the Hölder descriptor. If we set the maximum number of points to the original 129 used by the canonical descriptor, then we can propose a combinatorial search problem where the goal is to find the optimal subset of these points using as objective the performance criteria presented in Section 2.2. Notice that the problem stated in this way only considers performance on the matching tests, it does not explicitly search for the smallest number of dimensions. The search problem was posed in this way for the following reasons. First, we are interested in finding the best possible descriptor, one that achieves a performance that is comparable to the canonical Hölder descriptor. In fact, if the search

Parameter	Description and value
Representation	Binary string.
Population size	100.
Generations	100.
Selection	Fitness proportional.
Crossover	Mask crossover; $p_c = 0.9$ .
Mutation	Single bit mutation; $p_\mu = 0.1$ .
Survival	Elitism of the best 15%.
Training pairs	$N = 3$ .

Table 1: GA run-time parameters.

Name	Transformation	No. of Images
Nueva York	Rotation	35
Van Gogh	Rotation	17
Monet	Rotation	18
Graph	Illumination	12
Mosaic	Illumination	18

Table 2: Image sequences used to evaluate the performance of the H-GA descriptor

converges towards a descriptor of dimension  $\approx 129$ , then so be it. This would confirm that the original descriptor was the best possible construction given the proposed sampling grid. Second, if we add a second objective, namely the size of the descriptor, then we are faced with a multi-objective problem. However, from a practical perspective a Pareto front of solutions would only add another level of analysis that goes beyond the goal of this work.

#### 4.1 The genetic algorithm

The problem described above is a combinatorial search in 129 dimensions, exactly the kind of search problem in a high-dimensional space in which a GA thrives. Therefore, the task of choosing which points will be used to build the descriptive vector is assigned to a GA. Figure 4b shows the original sample points used by the Hölder descriptor, these act as the upper bound for the GA search. Therefore, the chromosome of each individual is expressed as a binary string  $B = (b_1, b_2, \dots, b_{129})$  of 129 bits. Each bit is associated with one of the admissible sample points. Hence, when a bit is set to 1 then the corresponding sample point is used to build the descriptor. Conversely, if a bit is zero then the sample point is not considered.

The fitness of each individual is based on the F-Measure, given in Equation 1. It is important to note that each  $F_\beta$  value depends on a single recall/1-precision pair, and that the recall/1-precision curves contain 20 such points. Therefore, the mean value of the F-Measure  $\bar{F}_\beta$  is used to characterize each performance curve. Moreover, if  $N$  pairs of images are used to train each individual, we then have the same number of recall/1-precision curves and corresponding  $\bar{F}_\beta$ s. The final fitness measure for each individual is posed for as a minimization task with the following cost function,

$$f(B) = -\frac{1}{N} \sum_{i=1}^N \bar{F}_\beta^i. \quad (6)$$

Note that the cost function does not explicitly favor individuals that produce smaller descriptors. The assumption is that evolution will be less constrained, and will tend to favor individuals that produce the best performance on the matching

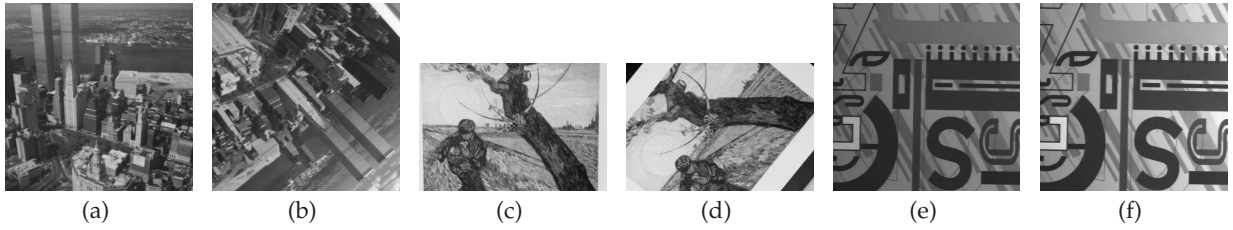


Figure 5: Training pairs of images used to assign fitness, each pair has a reference and a transformed image. (a,b) New York image with rotation transformation; (c,d) Van Gogh image with rotation; and (e,f) Graph image with illumination change.

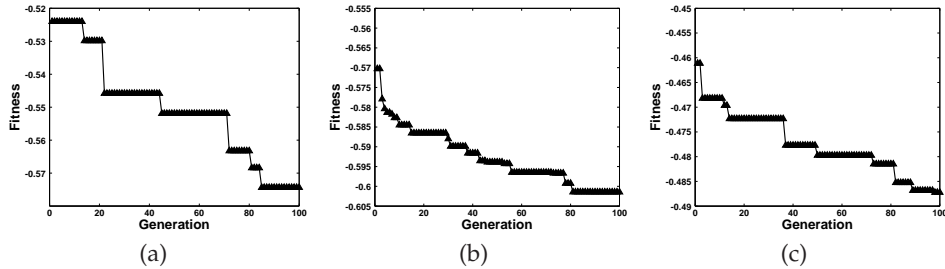


Figure 6: The convergence graphs of the three best experiments.

tests. The goal is to find the optimal subset of sample points. The GA was configured using the parameters shown in Table 1, from which we can see that the algorithm is basically standard. The fitness function uses three pairs of images for training, each pair is shown in Figure 5. Using this setup each run required between eight and ten hours of computation time, and a total of twenty runs of the algorithm were carried out in order to validate the performance of the proposal. Despite the long run-times, it is important to remember that this should be considered as a training step, and normal use of the final solutions does not require similar executions.

## 5. EXPERIMENTS AND RESULTS

In this section we present the best three results found using the proposed GA search. Figure 6 shows the convergence plots of the best fitness at each generation for each of the three best experiments, (a), (b) and (c). In all three cases we can see a steady and progressively improving convergence.

Then, in Figure 7 we show the phenotype of the best individual found in each of these runs. For comparison, the sample points chosen by each individual are marked using the same region presented earlier in Figure 4, and the total number of points is specified.

The first observation we can make is that even if all three of the solutions are different, they all construct a smaller vector than the canonical Hölder descriptor. For instance, the best solution from run (a) uses 75% of the original sample points, the best solution from run (b) uses only 31%, and the solution found in run (c) uses only 37% of the maximum number of sample points. In all three cases the reduction in descriptor size is significant, particularly for runs (b) and (c).

However, the reduction in size and the good fitness scores do not imply that the descriptors will achieve a high level of performance on a wider variety of test cases. Therefore, the performance of the evolved descriptors must be validated on more images and compared relative to the performance of the canonical Hölder descriptor. For such a comparison

we use similar criteria as those used in [13, 24], testing on different image pairs and testing over complete image sequences which contain a base image and a series of progressively transformed images. However, in order to simplify the following discussion, we only present the best solution found in all of the runs, which was the sampling pattern obtained in run (b), which we denote as  $H - GA$  (Hölder descriptor with Genetic Algorithm).

Before we compare with the canonical Hölder descriptor a few comments are necessary. First, solutions (a) and (c) were inferior to (b) based on their performance plots obtained with the matching tests. Therefore, we can say that the algorithm sometimes converged towards local-optima. Second, surprisingly the best performance was obtained using the solution that uses the least amount of sample points, which is strong evidence that suggests that the dimensions of the Hölder descriptor can be significantly reduced. Moreover, it is informative to see the spatial distribution of sample points suggested by the  $H - GA$  descriptor, see Figure 7b. The test region contains a typical corner structure, with the actual corner nearly corresponding with the center of the detected region. It is evident that most of the points are located very close to the center of the region, in the first two concentric rings. This result is consistent with the SIFT algorithm, where the contribution that each point has towards building the descriptor is inversely proportional with its distance to the central point. Thus, the descriptor is building a description of the local region that heavily relies on the appearance of the central part of the region. On the other hand, the points that are sampled in the final two rings, farthest away from the central point, seem to be distributed in an almost symmetric manner along a tangential line to the corner. The distribution of points does not seem to be arbitrary, because the organization is relative towards the principal direction of the gradient within the region. Moreover, most points appear almost entirely outside of the inner surface of the corner structure. This is consistent with the assumption that most corner structures will tend to

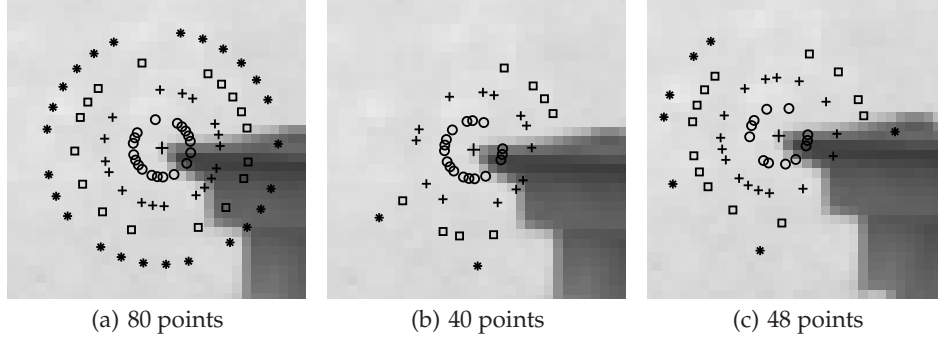


Figure 7: The phenotype of the best individual from each run. The figure shows the sample points that are used to build the descriptor.

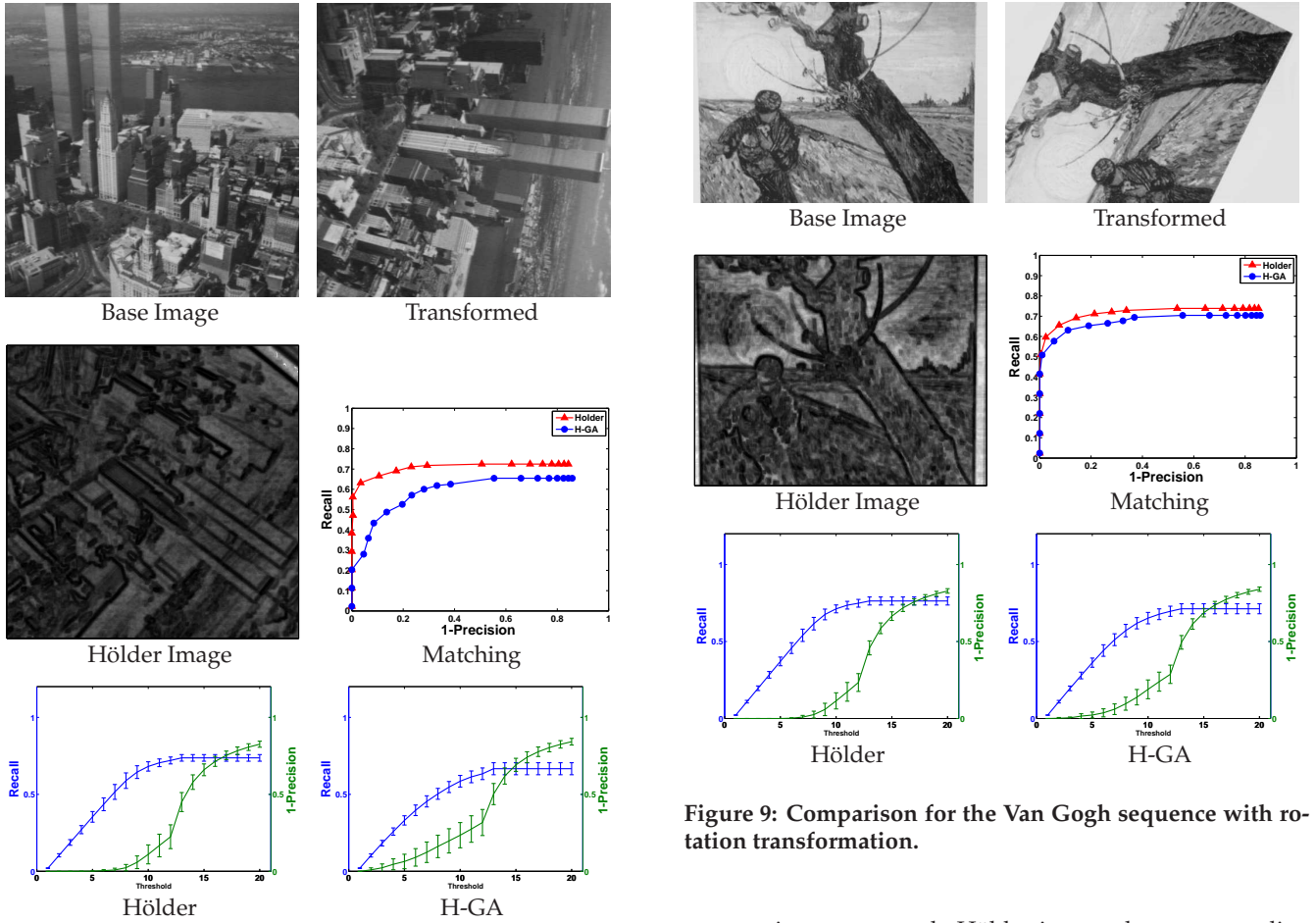


Figure 8: Comparison for the New York sequence with rotation transformation.

be homogeneous, and that discriminative information contained around a corner will not be within its inner flat structure.

Now, the comparison between  $H - GA$  and the canonical Hölder descriptor is presented in Figures 8, 9, 10, 11 and 12. In each test we compare the descriptors using a sequence of progressively transformed images and show the base image,

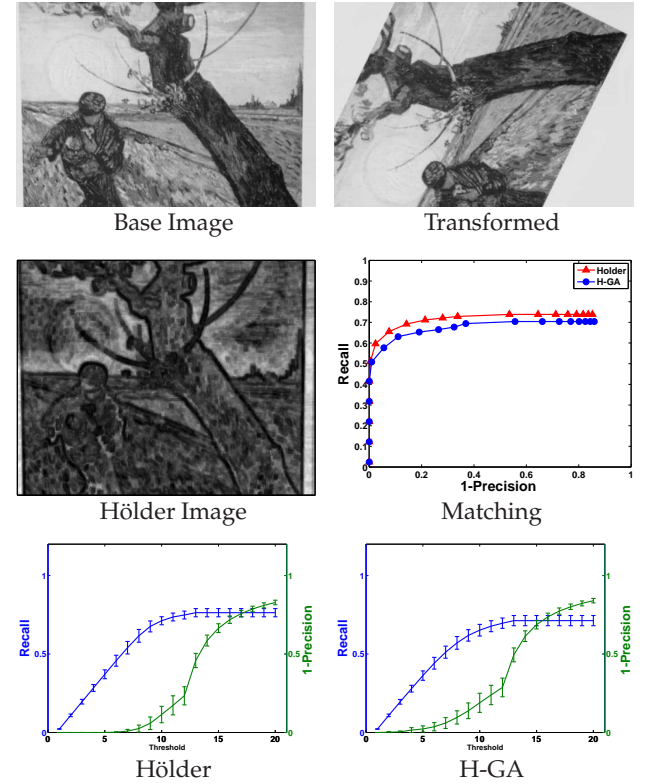


Figure 9: Comparison for the Van Gogh sequence with rotation transformation.

one test image, a sample Hölder image, the corresponding recall vs. 1-precision curve between the base and test image, and two plots that show the average performance of each descriptor computed for the complete test sequence; these last two plots require further explanation. The plots have a double y-axis that show the average recall and 1-precision scores computed for all of the images in each sequence, they also show the standard deviation for these measures. The x-axis in these plots corresponds with the different thresholds used for matching, see Section 2.2. In these plots an optimal performance is a horizontal recall curve close to one, and horizontal 1-precision curve close to zero. Table 2 summarizes the image sequences used to perform our experimental com-



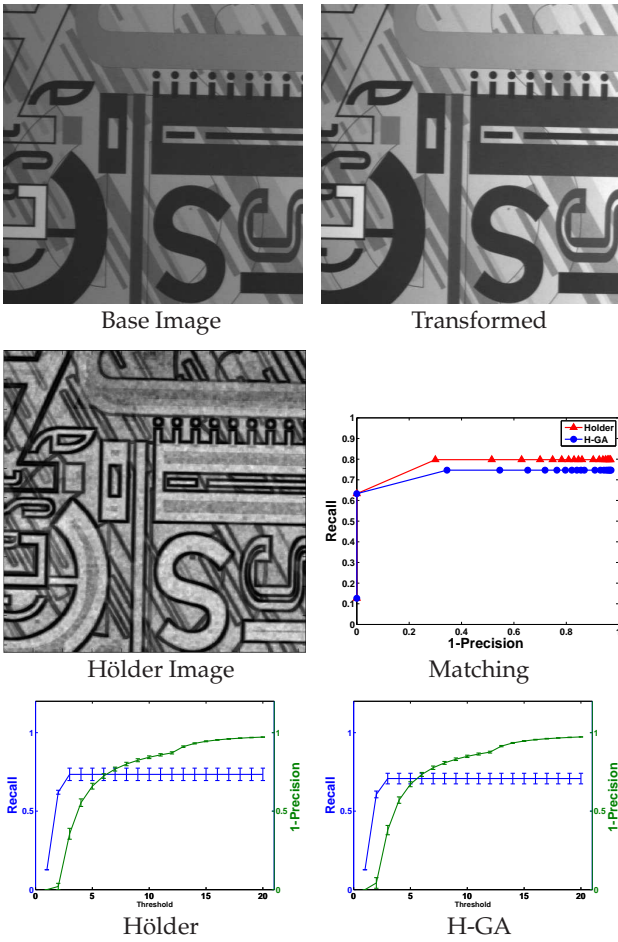


Figure 10: Comparison for the Graph image sequence with illumination change.

parisons, it gives the name of the sequence, the type of image transformation and the number of images in each sequence.

Figures 8, 9, 10 use the same image sequences from which the three training pairs were obtained. However, in all cases the test image that is used is different than the one that was used to compute the value of the cost functions, those are shown above in Figure 5. Several observations are pertinent here. First, we can appreciate that in some cases the performance between both descriptors is very similar, see Figures 9 and 10. Particularly, we can see that on average the performance is very similar over the complete sequences. In other cases, namely in Figure 8 and 11, the canonical descriptor is better. And still yet, in Figure 12 the H-GA achieves a better performance. However, as is obvious from previous comparative works [13] most of these differences should be negligible, and in the performance can be regarded as equivalent. Moreover, an important part of these comparisons is the manner in which the recall/1-precision curves behave, not just the level of recall that they reach [13]. Obviously, a perfect descriptor would achieve a recall equal to 1 for any precision, however this should not be expected in a real-world test. Therefore, for practical purposes what is desired is a horizontal curve that achieves a high (above 0.6) recall that remains steady. Conversely, a slowly increasing curve shows

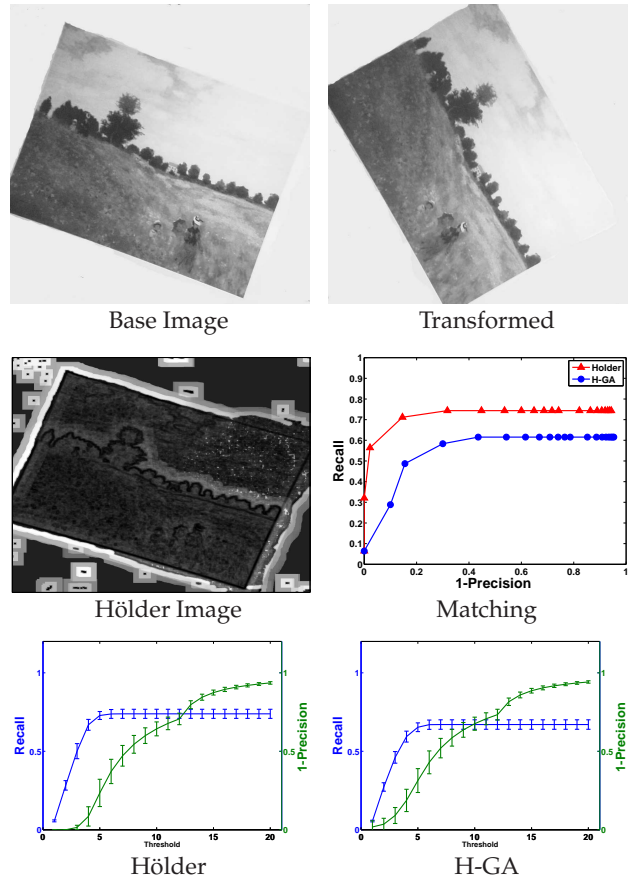


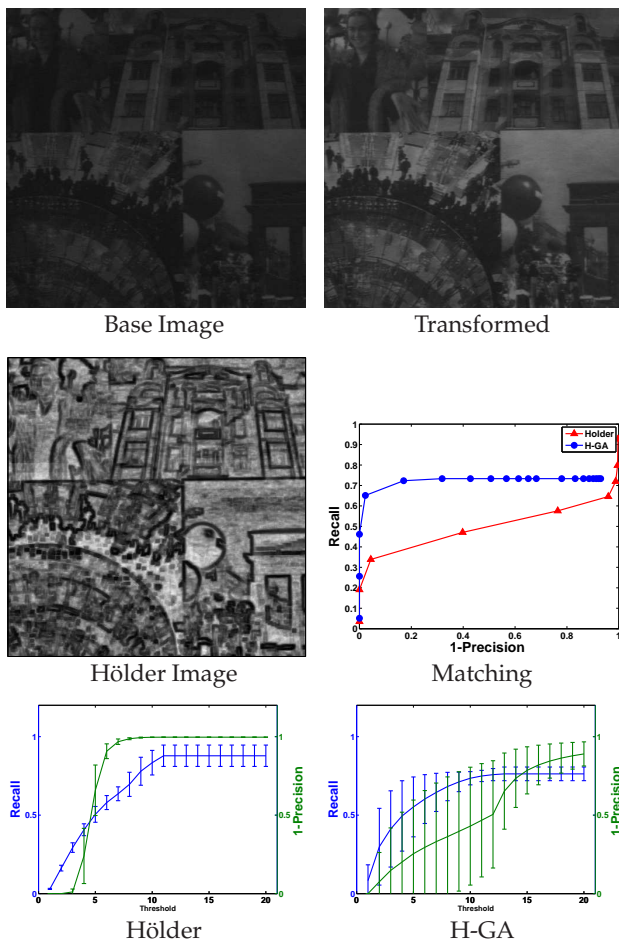
Figure 11: Comparison for the Monet image sequence with rotation transformation.

that the descriptor is affected by the image degradation induced by the transformation; i.e., the descriptor is less invariant. Under such considerations, we can see that in fact the performance of both descriptors is quite similar. However, H-GA achieves these performance scores using nearly 70% less information than the canonical descriptor.

## 6. SUMMARY AND CONCLUSIONS

This work addresses the problem of optimizing a descriptor for local image features. The study focuses on the Hölder descriptor because it achieves state-of-the-art performance, and because it relies on a very simple algorithm. The goal is to find the optimal set of sample points from which to compute the Hölder exponent and construct the Hölder descriptor. This task is posed as a combinatorial search problem and solved using a genetic algorithm. Fitness depends on the F-Measure of the descriptor computed on standard tests of region matching. The GA search produced several solutions that produce much more compact region descriptors. In fact, the best solution found by the GA, here called H-GA, uses only 31% of the dimensions from the canonical version of the descriptor and still achieves a similar performance. This suggests that the problem we have posed is multi-modal and that at least two optima exist, one of which is significantly more compact, and more efficient, than the other.





**Figure 12: Comparison for the Mosaic image sequence with illumination change.**

## References

- [1] J. Bala, K. D. Jong, J. Huang, H. Vafaie, and H. Wechsler. Using learning to facilitate the evolution of features for recognizing visual concepts. *Evol. Comput.*, 4(3):297–311, 1996.
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008.
- [3] M. Calonder, V. Lepetit, and P. Fua. Keypoint signatures for fast learning and recognition. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 58–71, Berlin, Heidelberg, 2008. Springer-Verlag.
- [4] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero. Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Computer Vision and Image Understanding*, 106(2-3):258–269, 2007.
- [5] S. Jaffard. Wavelet techniques in multifractal analysis. In *Fractal Geometry and Applications: A Jubilee of Benoit Mandelbrot, Proceedings of Symposia in Pure Mathematics*, volume 72, pages 91–151, 2004.
- [6] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Proceedings of CVPR*, volume 2, pages 506–513. IEEE Comp. Soc., 2004.
- [7] P. Legrand. *Debruitage et interpolation par analyse de la regularite Hölderienne. Application a la modelisation du frottement pneumatique-chaussee*. PhD thesis, Université de Nantes, France, 2004.
- [8] P. Legrand and J. Lévy-Véhel. Local regularity-based interpolation. In *WAVELET X, Part of SPIE's Symposium on Optical Science and Technology*, volume 5207, 2003.
- [9] P. Legrand and J. L. Vehel. Local regularity - based image denoising. *ICIP03, Spain, IEEE International Conference on Image Processing*, pages 377–380, September 14-17, 2003.
- [10] J. Lévy-Véhel. *Fractal Image Encoding and Analysis*, chapter Introduction to the Multifractal Analysis of Images, pages 299–341. 1998.
- [11] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of ICCV*, volume 2, pages 1150–1157. IEEE Computer Society, 1999.
- [12] S. Mallat. *A wavelet tour of signal processing*. Elsevier, San Diego, CA, 2nd edition, 1999.
- [13] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.
- [14] C. B. Perez and G. Olague. Evolutionary learning of local descriptor operators for object recognition. In *GECCO '09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 1051–1058, New York, NY, USA, 2009. ACM.
- [15] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, May 1997.
- [16] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.
- [17] W. Siedlecki and J. Sklansky. A note on genetic algorithms for large-scale feature selection. *Pattern Recogn. Lett.*, 10(5):335–347, 1989.
- [18] Z. Sun, G. Bebis, and R. Miller. Object detection using feature subset selection. *Pattern Recognition*, 37(11):2165–2176, 2004.
- [19] C. Tricot. *Curves and Fractal Dimension*. Springer-Verlag, 1995.
- [20] L. Trujillo and G. Olague. Synthesis of interest point detectors through genetic programming. In M. Cattolico, editor, *Proceedings of GECCO*, volume 1, pages 887–894. ACM, 2006.
- [21] L. Trujillo and G. Olague. Scale invariance for evolved interest operators. In *Proceedings of EvoIASP 2007*, LNCS, pages 423–430. Springer-Verlag, 2007.
- [22] L. Trujillo and G. Olague. Automated design of image operators that detect interest points. *Evolutionary Computation*, 16(4):483–507, 2008.
- [23] L. Trujillo, G. Olague, F. Fernández de Vega, and E. Lutton. Selecting local region descriptors with a genetic algorithm for real-world place recognition. In *Proceedings of EvoIASP 2008*, LNCS, pages 325–334, 2008.
- [24] L. Trujillo, G. Olague, P. Legrand, and E. Lutton. Regularity based descriptor computed from local image oscillations. *Optics Express*, 15:6140–6145, 2007.
- [25] L. Trujillo, G. Olague, E. Lutton, and F. Fernández de Vega. Multiobjective design of operators that detect points of interest in images. In M. Cattolico, editor, *Proceedings of GECCO*, pages 1299–1306, New York, NY, USA, 2008. ACM.
- [26] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Found. Trends Comput. Graph. Vis.*, 3(3):177–280, 2008.